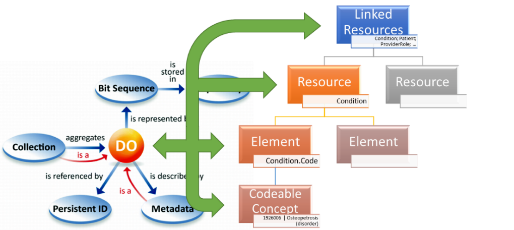
### The Problem

The concept of FAIR digital object is quite wide and can vary in term of granularity and type of data that should be represented.

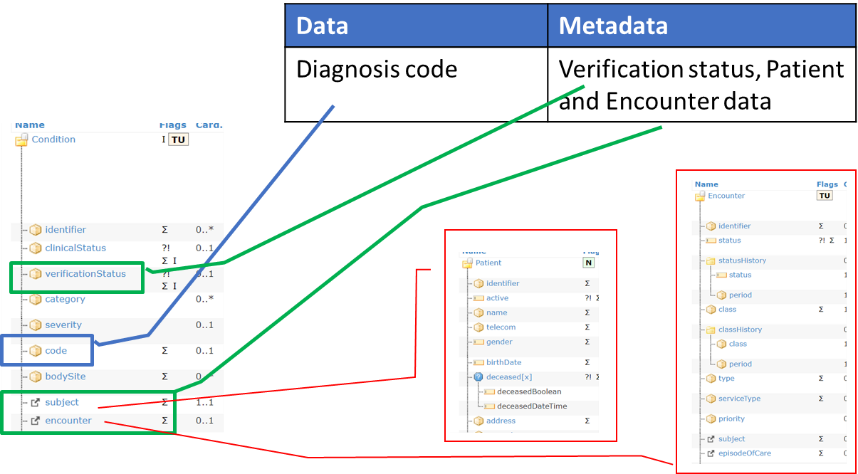
In fact, a FAIR digital object can be a single atomic information (e.g., a coded diagnosis) up to a collection of data (e.g., a data set). Data, moreover, can represent quite different kinds of information: it might be for example a waveform, an image, a condition, a medication, or other kinds of data.

This variety makes not so straightforward the mapping between FAIR data objects and the FHIR resources covering a large range of resources and elements (see figure below).



**Figure 1 – FAIR Digital Object representation**

Similar considerations can be done also for the metadata, where, depending on what the metadata is, the metadata elements can be represented within the same resource documenting the data, or by a set of linked resources.

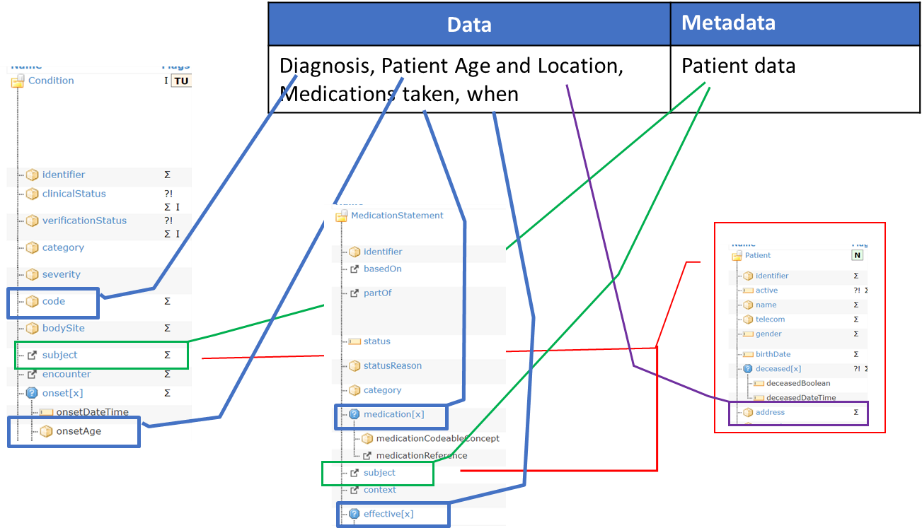


**Figure 2 – Mapping FAIR data and metadata**

#### Data vs metadata

In addition to the issue reported above, depending on what a "FAIR object" is the boundary between data and metadata may change. That means that based on the context, the same information can be considered as part of the data or of the metadata. This makes even more complex the mapping to the FHIR resources.

The figure below shows how information that were part of the metadata in the previous example are now part of the data; and how metadata and data elements are represented in the same FHIR resource.



**Figure 3 – Mapping FAIR data and metadata**

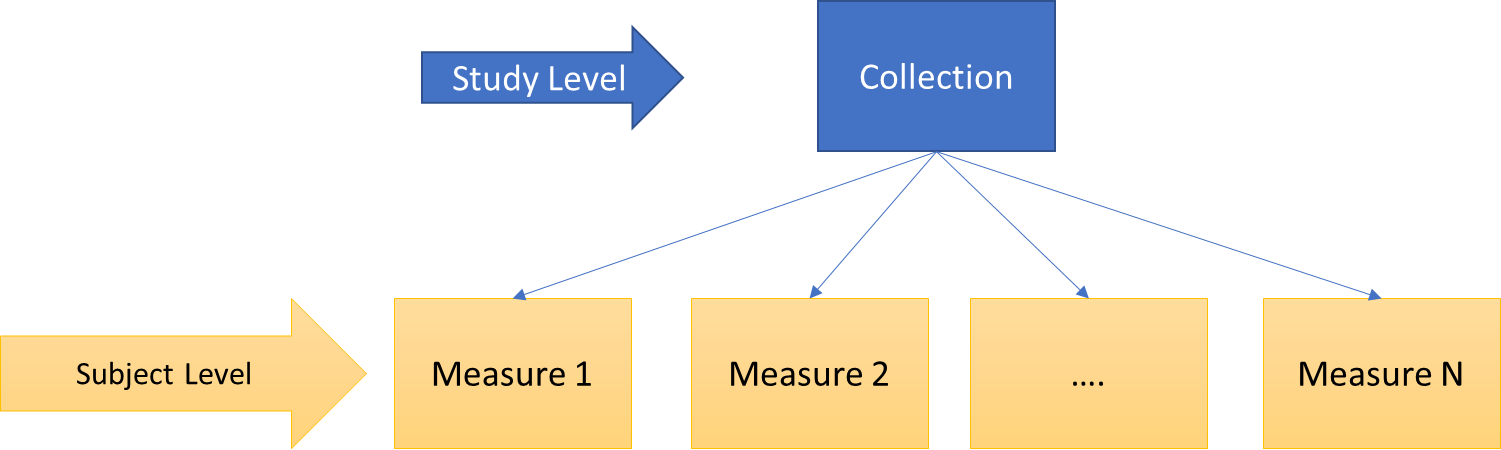
#### Metadata identification

In consideration of the previous arguments, the expectation of having a persistent and unique ID that identifies the metadata FAIR data object, distinct from the data FAIR data object, cannot be always satisfied in the FHIR space, except for specific contexts.

### The approach

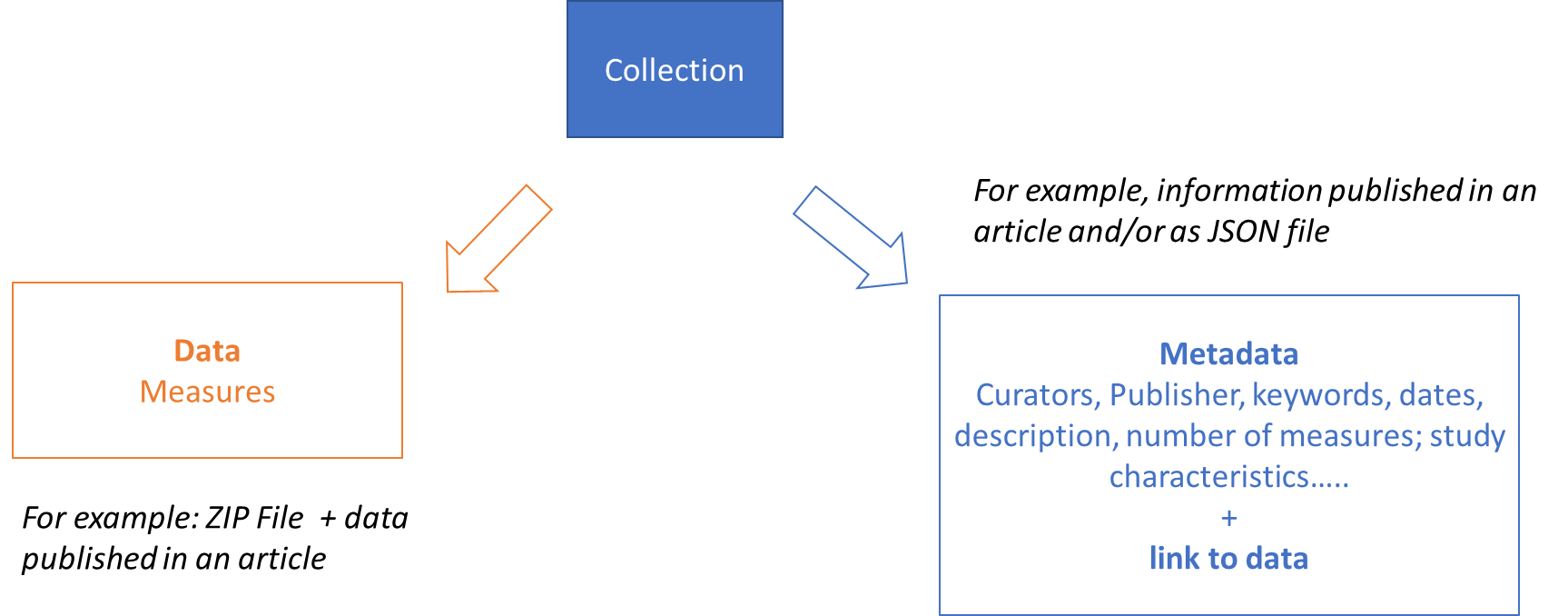
Even though many levels of granularity can be considered in this work we distinguish two main levels:

1. The **study level**: describing the collection of data referring to a specific study, publication, usage context.
2. The **subject Level**: providing the metadata/data information associated to a single subject.



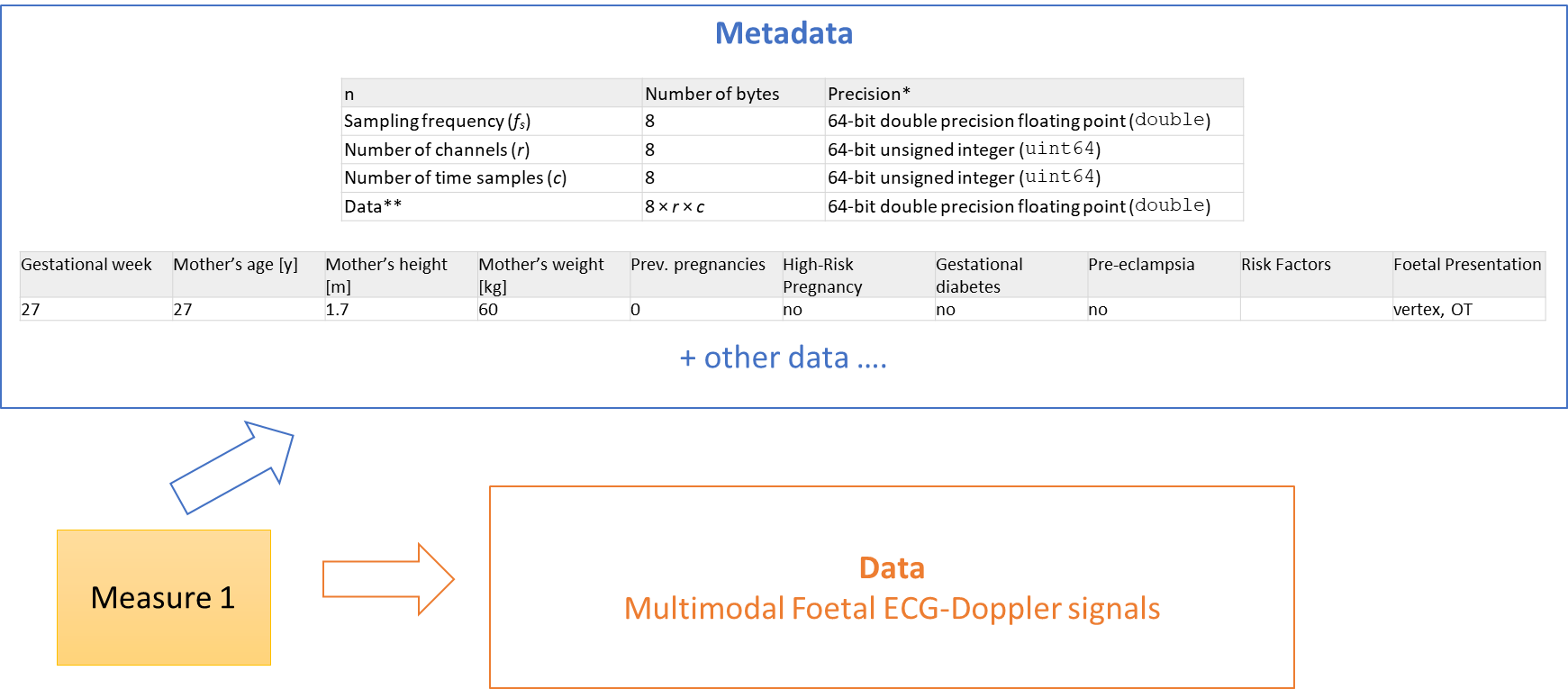
**Figure 4 - Level of data objects granularity considered in this guide**

If the boundary between metadata and data is quite clear for the collection (study level)



**Figure 2 – Metadata and data for a data collection.**

This might be not so true for the instance level, where information as the gestational week or the mother height and weight, might be considered metadata for the signals, but also patient level data.



**Figure 3 – Metadata and data at the subject level.**

### Metadata Representation in FHIR

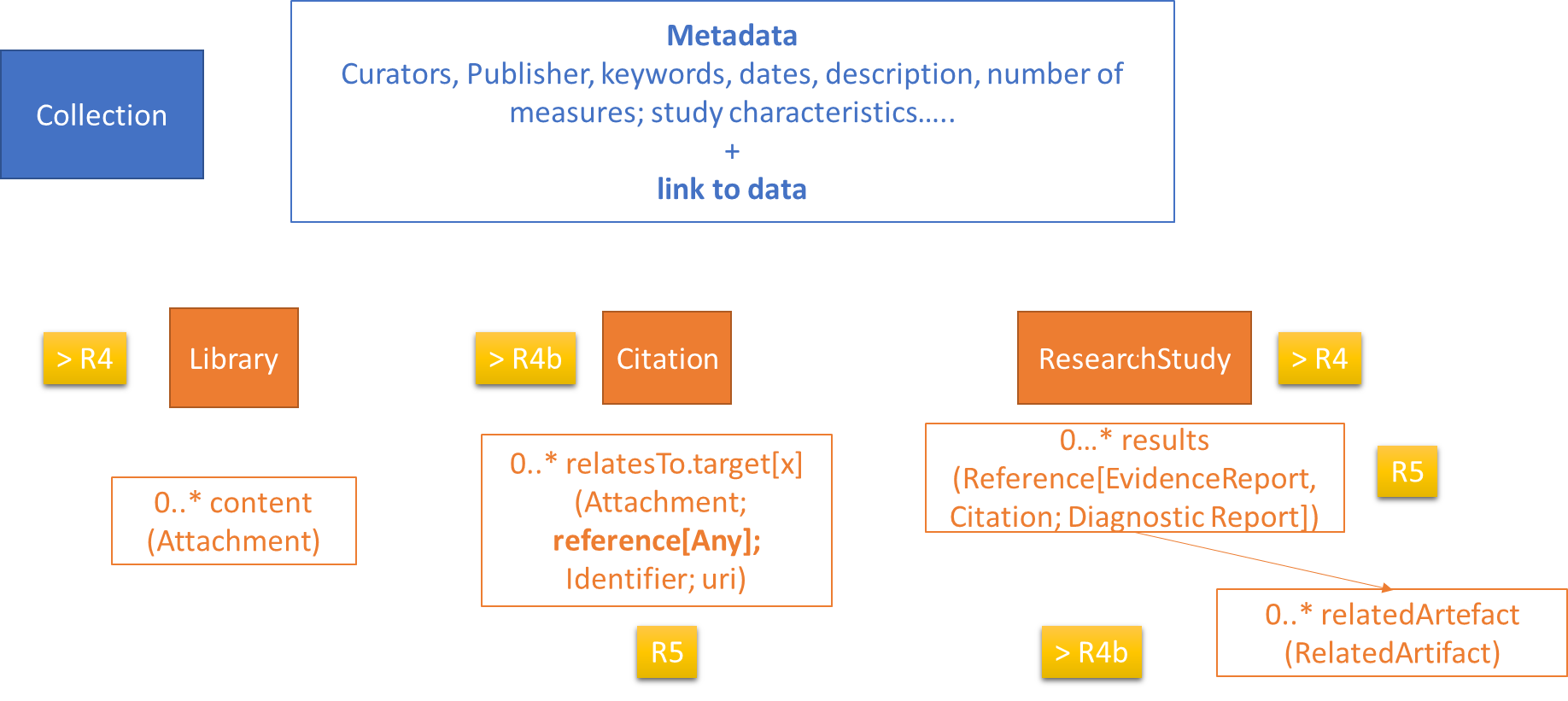
#### Study Level (Collection)

HL7 FHIR provides several candidates resources that could be used to represent collection of data (e.g., Bundle, Lists..); some of them however seem not to fully support the “rich metadata” FAIR requirement.

At this stage three FHIR resources have been identified to represent Collection metadata:

* Library
* Citation
* ResearchStudy

The choice of the resource depends on the usage context.



**Figure 4 – Data Collection Metadata representation**

The figure highlights the version of FHIR from which this resource is available and how the linkage between metadata and data is realized (with the FHIR version where this is derived from).

Considering three possible kinds of data objects:

1. Non-FHIR objects
2. FHIR and non-FHIR objects
3. FHIR-only objects

The following table highlights how this linkage could be realized.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Library content** | **Citation relatedTo.target[x]** | **Research Study results** |
| **Applicable to** | FHIR R4 | FHIR R5 (R4b) | FHIR R5 |
| **Non-FHIR binary** | link via uri | link via uri | EvidenceReport.relatedArtifact |
| **Combination of FHIR and non-FHIR objects** | link via uri  *(Add an extension ?)* | Reference to a FHIR resource | EvidenceReport.relatedArtifact |
| **Only FHIR** | link via uri  *(Add an extension ?)* | Reference to a FHIR resource | EvidenceReport.relatedArtifact |

#### Subject Level (e.g. Patient)

As describe above, at the subject level, the boundary between metadata and data is not always so sharp, since the classification can depend on the perspective: for example, the gestational age of a EGC measurement subject, is a measure metadata, but also a subject data.

This guide will not therefore to attempt to **prescribe any tight separation of metadata/data at the subject level.**

Concerning three possible kinds of data objects that have been mentioned above:

1. Non-FHIR objects
2. FHIR and non-FHIR objects
3. FHIR-only objects

In case of non-FHIR objects it should be evaluated feasibility and cost/benefit of transforming these objects (completely or partially) in FHIR; or having them referred by a subject level FHIR resource (see the group resource case)

For what concern the intermediate case (FHIR and non-FHIR objects) this should be evaluated case by case depending on where and how FHIR resources are used and the possibility to FHIRify the non-FHIR objects. Depending on this the solutions proposed for the other two cases should be assessed

In case of all the data represented by using FHIR resources, two main possible approaches for FAIR subject level data/metadata can be considered, leaving then to the implementers the choice about the most appropriate solution in a specific context:

1. Direct references: the collection refers directly the resources describing the (main) subject level data, relying on the existing FHIR resource relationships to describe the entire subject level set of data/metadata. For example, a study refers a Condition resource, using the *subject*, the *encounter* and the *evidence* references to document the patient, encounter and other information used as evidences.
2. Group resources. The second approach is to group all the subject level data, by using appropriate FHIR resources (EvidenceReport, Bundle[collection], Composition, List; DocumentManifest). This grouping resource may play a metadata role.

Open questions:

Should we leave all these options open or suggest one or few specific solutions?

*< expand this part explain pros and contra of the different options>*

*< to be continued>*

### Conclusions

*< move to another page?>*

* Do not pretend to address all the identified issues and realize all the FAIR recommendations since the beginning: **FAIRness is a continuous improvement process**.
* Follow a progressive approach focusing on the characteristics that are relevant for your context, balancing advantages and costs. For example, you may reasonably accept that metadata and data information at the patient level may be represented by a set of already existing FHIR resources, focusing your effort on a well identifiable and distinct representation of the metadata of the data collection at the study level.